An analysis of the effect of vitamin C on the tooth length growth of guinea pigs.

Guillaume Dreyer

Analysis report.

We study the data set 'ToothGrowth', and draw some conclusions from it.

1. Load the 'ToothGrowth' dataset.hist

```
data(ToothGrowth)        # load the 'ToothGrowth' dataset
dim(ToothGrowth)         # dimension of the data set
```

```
## [1] 60  3
```

```
head(ToothGrowth)        # visualize the first rows
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

2. Basic summary.

Origin of the dataset.

The response is the **length of odontoblasts** (teeth) in each of 10 guinea pigs at each of three **dose levels of Vitamin C** (0.5, 1, and 2 mg) with each of two delivery methods (**orange juice** or **ascorbic acid**).

Description of the dataset.

The 'ToothGrowth' dataset contains 60 observations on 3 variables.

1. **'len'**: a numeric variable indicating the tooth length.
2. **'supp'**: a factor variable that indicates how the Vitamin C dose was administrated: 'VC' for ascorbic acide, 'OJ' for orange juice.
3. **'dose'**: a numeric variable indicating the administrated dose in milligrams.

3. Exploritary analysis.

Compare the administration methods.

We begin with creating a data frame **"meanGrowth"** that shows the average tooth length within the two groups: **'VC'** (ascorbic acide) or **'OJ'** (orange juice). The code can be find in the **Appendix**.

```
meanGrowth
```

```
##     len supp dose
## 1 13.23   OJ  0.5
## 2 22.70   OJ  1.0
## 3 26.06   OJ  2.0
## 4  7.98   VC  0.5
## 5 16.77   VC  1.0
## 6 26.14   VC  2.0
```

The above data frame suggests that the tooth lentgh growth is higher for the guinea pigs in the 'OJ' group than for those in the 'VC' group. One can evaluate this hypothesis by testing it against the **null hypothesis**, namely, "there is a zero mean difference between the two groups". Let's run a paired **t test** using the *t.test* command!

```
TTest$conf.int; TTest$statistic
```

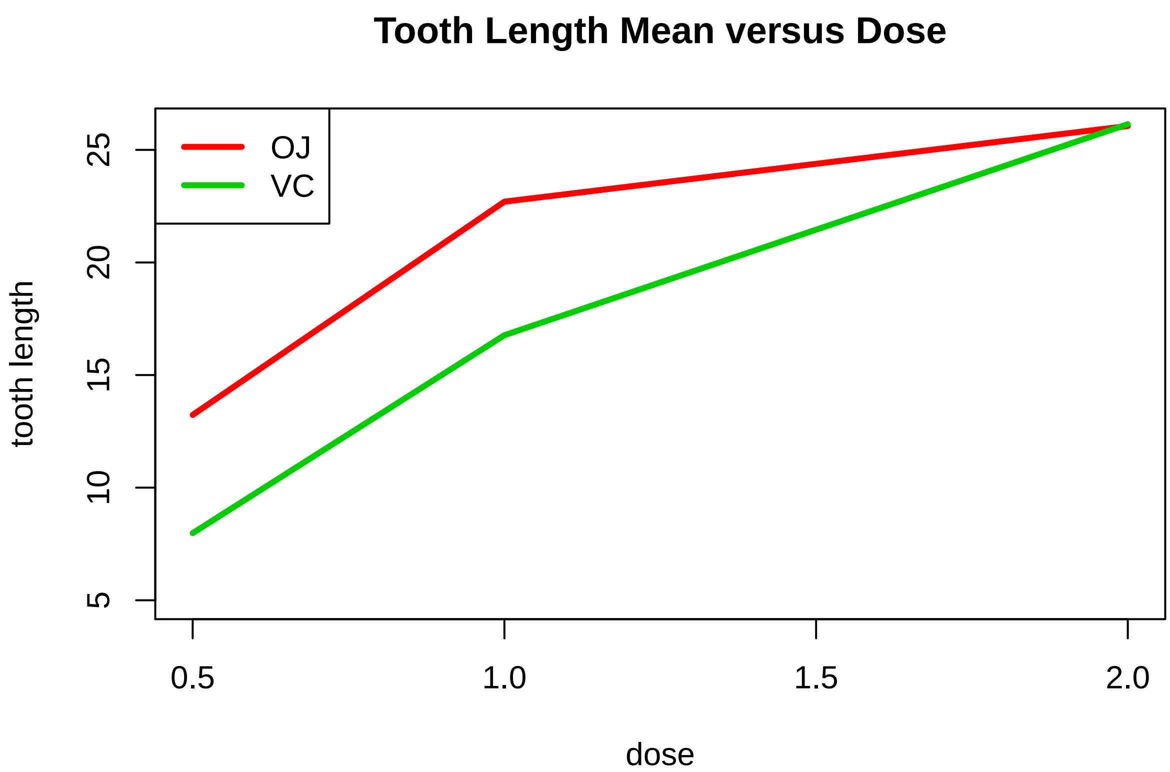
```
## [1] 1.408659 5.991341
## attr(,"conf.level")
## [1] 0.95
```

```
##        t
## 3.302585
```

The **confidence interval** 1.4086586, 5.9913414 for the mean difference **does not contain 0**. Moreover, the **t statistic** 3.302585 is larger than 2. We thus **reject** the null hypothesis in favor of the **alternative hypothesis**.

However, observe from the data set **"meanGrowth"** that, if a clear difference in the tooth growth exists between the two delivery methods for doses equal to 0.5 or 1.0 mg, this tends to disappear for a dose of 2.0 mg, as shown in the graph below.

```
x <- subset(meanGrowth, supp == 'VC')[['dose']]
y_OJ <- subset(meanGrowth, supp == 'OJ')[['len']]
y_VC <- subset(meanGrowth, supp == 'VC')[['len']]
graph(x,y_OJ,y_VC)
```



One can test this hypothesis by comparing the difference between the two groups for equal doses. Let's run a series of t **tests**...

```
TTest1$conf.int; TTest1$statistic
```

```
## [1] 1.263458 9.236542
## attr(,"conf.level")
## [1] 0.95

##        t
## 2.979105
```

```
TTest2$conf.int; TTest2$statistic
```

```
## [1] 1.951911 9.908089
## attr(,"conf.level")
## [1] 0.95

##       t
## 3.37212
```

```
TTest3$conf.int; TTest3$statistic
```

```
## [1] -4.328976  4.168976
## attr(,"conf.level")
## [1] 0.95
```

```
##           t
## -0.04259204
```

The **confidence intervals** for the mean difference, for doses equal to 0.5 and 1mg are 1.2634583, 9.2365417, and 1.9519109, 9.9080891, respectively. Moreover, the associated **t statistic** are 2.9791047, and 3.3721195, respectively. We therefore **reject** the **null hypothesis** for 0.5 and 1 mg doses in favor of the **alternative hypothesis**. However, for a 2mg dose, one **fails to conclude**. Indeed the confidence interval -4.3289765, 4.1689765 contains 0, but the associated t statistic -0.042592 is too small.

4. Conclusion

Our analysis demontrates a clear difference in the tooth length growth for the two delivery methods for doses less or equal to 1mg. However, this phenomenon tends to attenuate for larger doses.

Appendix.

```r
# create a data set "meanGrowth" that shows the mean tooth length
# for a given dose and a delivery method
DF <- ToothGrowth
meanGrowth <- data.frame()
groups <- split(DF, list(DF$supp, DF$dose))
for (i in 1:length(groups)){
        row <- data.frame('len' = colMeans(groups[[i]][1]),
                          'supp'= groups[[i]][1,2],
                          'dose'= colMeans(groups[[i]][3]))
        meanGrowth <- rbind(meanGrowth, row)
        meanGrowth <- meanGrowth[order(meanGrowth$supp),]
        rownames(meanGrowth) <- NULL
}
```

```r
# Confidence interval for the mean difference between the two groups
g2 <- subset(ToothGrowth, supp == 'OJ')$len
g1 <- subset(ToothGrowth, supp == 'VC')$len
TTest <- t.test(g2, g1, paired = TRUE, var.equal = FALSE)
```

```r
# the 'graph' function plots a single for two different groups
par(mfrow=c(1,1))
graph <- function(x,y1, y2){
        plot(x,y1,type="l",lwd = 3,
             main = "Tooth Length Mean versus Dose",
             ylim = c(5,26), xlab='dose', ylab='tooth length',
             col="red")
        lines(x,y2,col="green3",lwd = 3)
        legend("topleft",
               legend = c("OJ" ,"VC"),
               lty=c(1,1),lwd=c(3,3),
               col = c("red" ,"green3"), cex=1)
}
```

```r
# Confidence intervals for the mean difference between the two groups
# for a given dose.
dose_groups <- split(ToothGrowth, ToothGrowth$dose)
h1 <- subset(dose_groups[['0.5']], supp == 'VC')$len
h2 <- subset(dose_groups[['0.5']], supp == 'OJ')$len
hh1 <- subset(dose_groups[['1']], supp == 'VC')$len
hh2 <- subset(dose_groups[['1']], supp == 'OJ')$len
hhh1 <- subset(dose_groups[['2']], supp == 'VC')$len
hhh2 <- subset(dose_groups[['2']], supp == 'OJ')$len
TTest1 <- t.test(h2, h1, paired = TRUE, var.equal = FALSE)
TTest2 <- t.test(hh2, hh1, paired = TRUE, var.equal = FALSE)
TTest3 <- t.test(hhh2, hhh1, paired = TRUE, var.equal = FALSE)
```